# SWABEY OGILVY RENAULT

S.E.N.C.

PATENT AND TRADE MARK AGENTS

02DO
#2

May 9, 2000

**REGISTERED MAIL**

**Box NON-FEE**
**(Priority Document)**
Assistant Commissioner for Patents
Washington, D.C. 20231
U. S. A.

Sir:

**Re: United States Patent Application**

| | |
|---|---|
| **Title:** | **HIGH-CAPACITY WDM-TDM PACKET SWITCH** |
| **Serial No.:** | **09/550,489** |
| **Filed:** | **April 17, 2000** |
| **Inventor:** | **Maged E. BESHAI** |
| **Assignee:** | **Nortel Networks Corporation** |
| **Our File:** | **9-13528-61US MRW:lf** |

In support of the priority claim in the above-identified patent application, we enclose the required certified copy of corresponding Canadian Application No. 2,283,627.

Respectfully submitted,

Max R. Wood
Reg. No. 40,388
Agent of Record

MRW:lf
Encl.



MONTRÉAL OTTAWA TORONTO QUÉBEC
1981 MCGILL COLLEGE AVENUE, SUITE 1600, MONTRÉAL, QUÉBEC, CANADA H3A 2Y3
TELEPHONE (514) 845-7126 FACSIMILE (514) 288-8389

Office de la propriété intellectuelle du Canada

Un organisme d'Industrie Canada

Canadian Intellectual Property Office

An Agency of Industry Canada



#2
7/12/00
Jbebbs

# Bureau canadien des brevets

## Certification

# Canadian Patent Office

## Certification

La présente atteste que les documents ci-joints, dont la liste figure ci-dessous, sont des copies authentiques des documents déposés au Bureau des brevets.

This is to certify that the documents attached hereto and identified below are true copies of the documents on file in the Patent Office.

Specification and Drawings, as originally filed, with Application for Patent Serial No: **2,283,627,** on September 27, 1999, by **NORTEL NETWORKS CORPORATION,** assignee of Maged E. Beshai, for "High-Capacity WDM-TDM Packet Switch".

L. Régimbal

Agent certificateur/Certifying Officer

April 14, 2000

Date

Canada

(CIPO 68)

OPIC CIPO

## ABSTRACT OF THE DISCLOSURE

A self-configuring distributed packet switch which operates in wavelength division multiplexed (WDM) and time division multiplexed (TDM) modes is described. The switch comprises a distributed distributed channel switch core, the core modules being respectively connected by a plurality of channels to a plurality of high-capacity packet switch edge modules. Each core module operates independently to schedule paths between edge modules, and reconfigures the paths in response to dynamic changes in data traffic loads reported by the edge modules. Reconfiguration timing between the packet switch modules and the channel switch core modules is performed to keep reconfiguration guard time minimized. The advantage is a high-capacity, load-adaptive, self-configuring switch that can be distributed to serve a large geographical area and can be scaled to hundreds of Tera bits per second to support applications that require very high bandwidth and a guaranteed quality of service.

# HIGH-CAPACITY WDM-TDM PACKET SWITCH

## TECHNICAL FIELD

This invention relates generally to the field of data packet switching and, in particular, to a distributed very high-capacity switch having edge modules that operate in packet switching mode and core modules that operate in circuit switching mode, the core modules switching payload traffic between the edge modules using wavelength division multiplexing (WDM) and time division multiplexing (TDM).

## BACKGROUND OF THE INVENTION

Introduction of the Internet to the general public and the exponential increase in its use has focused attention on high speed backbone networks and switches capable of delivering large volumes of data at very high rates. In addition to the demand for higher transfer rates, many service applications are being developed, or are contemplated, which require guaranteed grade of service and data delivery at guaranteed quality of service. To date, efforts to grow the capacity of the Internet have largely been focused on expanding the capacity and improving the performance of legacy network structures and protocols. Many of the legacy network structures are, however, difficult to scale into very high-capacity networks. In addition, many legacy network protocols do not provide grade of service or quality of service guarantees.

Nonetheless, high capacity switches are known in the prior art. Prior art high capacity switches are commonly constructed as a multi-stage, usually three-

stage, architecture in which ingress modules communicate with egress modules through a switch core stage. The transfer of data from the ingress modules to the egress modules must be carefully coordinated to prevent contention and to maximize the throughput of the switch. Within the switch, the control may be distributed or centralized. A centralized controller must receive traffic state information from each of the ingress modules. Each ingress module reports the volume of waiting traffic destined to each of the egress modules. The centralized controller therefore receives traffic information related to traffic volume from each of the ingress modules. If, in addition, the controller is made aware of the class of service distinctions among the waiting traffic, the amount of traffic information increases proportionally. Increasing the amount of traffic information increases the number of control variables and results in increasing the computational effort required to allocate the ingress/egress capacity and to schedule its usage. Consequently, it is desirable to keep the centralized controller unaware of the class of service distinctions while providing a means of taking the class of service distinctions into account during the ingress/egress transfer control process.

This is accomplished in a rate-controlled multi-class high-capacity packet switch described in Applicant's copending United States Patent Application No. 09/244,284 which was filed on February 4, 1999. Although the switch described in this patent application is adapted to switch variable sized packets at very high speeds while providing grade-of-service and quality-of-service control, there still exists a need for a distributed switch that can form the core of a powerful

high-capacity, high-performance network that is adapted to provide wide geographical coverage with end-to-end capacity that scales to hundreds of Tera bits per second (Tbs), while providing grade of service and quality of service controls.

A further challenge in providing a powerful high-capacity, high-performance switch with wide geographical coverage is maintaining network efficiency in the face of constantly fluctuating traffic volumes. In response to this challenge, the Applicant also invented a self-configuring data switch comprising a number of electronic switching modules interconnected by a single-stage channel switch that includes a number parallel space switches, each having input ports and output ports. This switch architecture is described in Applicant's copending United States Patent Application entitled SELF-CONFIGURING DISTRIBUTED SWITCH which was filed on April 6, 1999 and assigned Application Serial No. 09/286,431. Each of the electronic modules is capable of switching variable-sized packets and is connected to the set of parallel space switches by a number of optical channels, each of the optical channels being a single wavelength in a multiple wavelength fiber link. The channel switching core permits any two modules to be connected by an integer number of channels. In order to enable the switching of traffic at arbitrary transfer rates, the inter-module connection pattern is changed in response to fluctuations in data traffic load. However, given the speed of optical switching equipment and the granularity of the channels, it is not always possible to adaptively modify the paths between modules to accommodate all data traffic variations. Consequently, it sometimes proves uneconomical to

establish under-utilized paths for node pairs with low traffic volumes. To overcome this difficulty, a portion of the data traffic flowing between a source module and a sink module is switched through one or more intermediate nodes. Thus, in effect, the switch functions as a hybrid of a channel switch and linked buffer data switch, benefiting from the elastic path capacity of the channel switch.

A concentration of switching capacity in one location is, however, undesirable for reasons of security and economics. The self-configuring distributed switch with a high capacity optical core described in Applicant's co-pending Patent Application is limited in capacity and limited to switching entire channels. Consequently, it is desirable to provide a high-capacity switch with a distributed core. Such a core has the advantages of being less vulnerable to destruction in the event of a natural disaster, for example. It is also more economical because strategic placement of distributed core modules reduces link lengths, requires less concentration of infrastructure and provides shorter paths for localized data traffic.

There therefore exists a need for a very high-capacity packet switch with a distributed core that is adapted to provide grade of service and quality of service guarantees. There also exists a need for a very high-capacity packet switch that provides intra-switch data paths of a finer granularity to reduce or eliminate a requirement for tandem switching.

**SUMMARY OF THE INVENTION**

It is an object of the invention to provide a very high-capacity packet switch with a distributed core

that is adapted to provide guaranteed grade of service and quality of service, as well as providing intra-switch data paths with a granularity that reduces or eliminates a requirement for tandem switching.

The invention therefore provides a high capacity packet switch that includes a plurality of core modules that operate in a circuit switching mode, and a plurality of edge modules that are connected to subtending packet sources and subtending packet sinks, each of the edge modules operating in a packet switching mode. The core modules switch payload traffic between the edge modules using wavelength division multiplexing (WDM) and time division multiplexing (TDM).

Each of the core modules is preferably a space switch. Any of the well known textbook designs for a space switch can be used. However, the preferred space switch is an electronic single-stage rotator switch, because of its simple architecture, ease of control and scalability. A one of the edge modules is preferably co-located with each core module and serves as a controller for the core module.

Each of the edge modules has a plurality of ingress ports and a plurality of egress ports. Each of the ingress ports has an associated ingress queue. An ingress scheduler sorts packets arriving in the ingress queues from the subtending packet sources, the sort being by destination edge module from which the respective packets are to egress from the high capacity packet switch for delivery to the subtending packet sinks. The ingress scheduler periodically determines a number of packets waiting in the ingress queues for each other respective edge module, and sends a payload traffic allocation vector to each of the controllers of the core

modules. The traffic allocation vector sent to a given controller relates only to a group of channels that connect the edge module to the given core module.

Each edge module also maintains a vector of pointers to the sorted payload packets, the vector of pointers being arranged in egress edge module order. A scheduling matrix for each slot in a time frame and each egress edge module is associated with the vector of pointers and determines a data transfer schedule for the ingress edge module.

Each ingress edge module also maintains an array of reconfiguration timing circuits, a one of the reconfiguration timing circuits being associated with each of the core modules. The reconfiguration timing circuits are respectively synchronized with time clocks in the respective edge modules that serve as controllers for the core modules. The reconfiguration timing circuits enable reconfiguration of channel switching in the core modules using a short guard time.

Each core module preferably comprises a plurality of rotator switches. Each rotator switch preferably accommodates a number of input channels equal to the number of edge modules, as well as a number of output channels equal to the number of edge modules. In a folded edge module configuration, each edge module preferably has one channel connected to an input port and one channel connected to an output port of each rotator switch. In an unfolded edge module configuration, each edge module is either an ingress module or an egress module. The ingress and egress modules are preferably arranged in co-located pairs. In the unfolded configuration, each ingress edge module preferably has one channel connected to an input port of each rotator

switch. Each egress module likewise preferably has one channel connected to an output port of each rotator switch.

The invention also provides a method of switching payload data packets through a distributed data packet switch. In accordance with the method, payload data packets are received from a subtending source at an ingress edge module of the distributed data packets switch. An identity of an egress edge module from which the data packets should egress from the distributed data packet switch is then determined. Using the identity of the egress edge module, the data packets are arranged in a sorted order with other data packets received so that the data packets are in a sorted order corresponding to the identity of the edge module from which the data packet should egress from the distributed data packet switch. The sorted data packets are transferred in fixed-length data blocks are switched through the core module to the egress module. The fixed-length data blocks contain concatenated packets of variable length, and the respective egress module parses the variable size packets according to methods known in the art. The fixed-length data blocks are switched through the core module to the egress module. Thereafter, the payload data packet is transferred to a subtending sink.

**BRIEF DESCRIPTION OF THE DRAWINGS**

The invention will now be explained by way of example only, and with reference to the following drawings, in which:

FIG. 1 is a schematic diagram of a high capacity WDM-TDM packet switch in accordance with the invention having a centralized core;

FIG. 2 is a schematic diagram of the high capacity WDM-TDM packet switch shown in FIG. 1 wherein the space switches in the core are single-stage rotator switches;

FIG. 3 is a schematic diagram of a high capacity WDM-TDM packet switch in accordance with the invention with a distributed core;

FIG. 4 is a schematic diagram of a high capacity WDM-TDM packet switch in accordance with the invention showing an exemplary distribution of the core modules and edge modules;

FIG. 5 is a schematic diagram of a data structure used in each edge module to facilitate a process of computing capacity request vectors in the edge modules;

FIG. 6 is a schematic diagram of a table used by an ingress edge module to determine a preferred core module for a connection to an egress module;

FIG. 7 is a schematic diagram of data structures used in each core module for capacity scheduling using capacity request vectors received from the edge modules;

FIG. 8 is a schematic diagram illustrating space switch occupancy in a four core-module distributed switch in which a matching method employing a packing-search discipline is used; and

FIG. 9 is a schematic diagram of data structures used to control the transfer of data blocks from an ingress module to core modules of a high capacity WDM-TDM packet switch in accordance with the invention.

It should be noted that throughout the appended drawings, like features are identified by like reference numerals.

**DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT**

FIG. 1 is a schematic diagram of a high-capacity WDM-TDM packet switch in accordance with the invention, generally indicated by reference 20. The packet switch 20 includes a plurality of edge modules 22, 24 shown for clarity of illustration in an "unfolded" configuration. In the unfolded configuration shown in FIG. 1, ingress edge modules 22 and egress edge modules 24 are separate switching modules constructed, for example, as described in Applicant's copending Patent Application Serial No. 09/244,824 which was filed February 4, 1999 and entitled RATE-CONTROLLED MULTI-CLASS HIGH-CAPACITY PACKET SWITCH, the specification of which is incorporated herein by reference. In a folded switch configuration, the ingress edge modules 22 and the egress edge modules 24 are combined into integrated switch modules of one ingress module and one egress module each, each integrated module having as many data ports as a sum of the data ports of the ingress edge module 22 and the egress edge module 24.

Located between the edge module pairs 22, 24 are a plurality of space switches 26 which serve as centralized core modules for the WDM-TDM packet switch 20. For the sake of scalability and switching speed, the space switches 26 are preferably electronic space switches, although optical space switches could be used and may become preferred when optical switching speeds improve. The space switches 26 are arranged in parallel and, as will be described below, are preferably distributed in collocated groups. The number of edge modules 22, 24 and the number of space switches 26 included in the WDM-TDM packet switch 20 is dependent on

the switching capacity required. In the example shown in FIG. 1, there are 256 (numbered as 0-255) ingress edge modules 22 and 256 egress edge modules 24. Each edge module 22 has egress ports to support 128 channels. In a typical WDM multiplexer, 16 wavelengths are supported on a link. Each wavelength constitutes a channel. Consequently, the 128 channels can be supported by eight optical fibers if cross-connectors are used, as will be explained below with reference to FIG. 3.

In order to ensure that any edge module 22 is enabled to send all of its payload traffic to any edge module 24, if so desired, each space switch 26 preferably supports one input channel for each module 22 and one output channel for each module 24. Therefore, in the example shown in FIG. 1, each space switch preferably supports 256 input channels and 256 output channels. The number of space switches 26 is preferably equal to the number of channels supported by each edge module 22, 24. In the example shown in FIG. 1, there are preferably 128 space switches 26, the number of space switches being equal to the number of channels from each ingress module 22.

FIG. 2 is a schematic diagram of a preferred embodiment of the WDM-TDM packet switch shown in FIG. 1. In accordance with a preferred embodiment, each of the space switches 26 is a single-stage rotator-based switch. In the rotator-based switch architecture, a space switch core is implemented as a bank of independent memories 28 that connect to the edge modules 22 of the switch through an ingress rotator 30. Traffic is transferred to the edge modules 24 of the switch 20 through an egress rotator 32. The two rotators 30, 32 are synchronized. A detailed description of the rotator switch architecture

is provided in United States Patent No. 5,745,486 that issued to Beshai et al. on April 28, 1998, the specification of which is incorporated herein by reference. In other respects, the switch architecture shown in FIG. 2 is identical to that shown in FIG. 1.

In the rotator switches 26, each bank of independent memories 28 is divided into a plurality of memory sections arranged in series. Each memory is preferably arranged in columns that are 128 bytes wide. Each memory is divided into a number of partitions, the number of partitions being equal to the number of egress edge module 24. The size of the memory portion governs a size of data block switched by the channel switching core. The size of the data block is a matter of design choice, but is preferably about 4-16 kilobits.

**Partitioning the Core**

The channel switching core is preferably partitioned into core modules and distributed for two principal reasons: economics and security. FIG. 3 is a schematic diagram of a preferred embodiment of a distributed WDM-TDM packet switch in accordance with the invention. A plurality of core modules 34 are geographically distributed. A plurality of cross-connectors 36, which may be, for example, very slow optical switches, connect a plurality of ingress and egress edge modules 22, 24 to the core modules 34. The cross-connectors 36 serve as multiplexers and thereby reduce the number of physical links required to connect each ingress and egress edge module 22, 24 to each core module 34. The core modules 34 preferably include an equal number of rotator switches. A WDM-TDM packet switch 20 of a size shown in FIGs. 1 and 2, with eight

core modules 34, includes 16 rotator switches 28 in each core module 34 when geographically distributed as shown in FIG. 3 (8 x 16= 128). If the ingress and egress edge modules 22, 24 are grouped in clusters of eight per cross-connector 36, then 64 cross-connectors are required to connect the ingress and egress edge modules 22, 24 to the core modules 34. The clustering of the ingress and egress edge modules 22, 24 and the number of cross-connectors 36 used in any given installation is dependent on network design principles well understood in the art and does not require further explanation. In any distributed deployment of the WDM-TDM packet switches, it is preferred that each ingress and egress edge module 22, 24 be connected to each space switch 26 of each core module 34 by at least one channel. The switch may be partitioned and distributed as desired with the exception that one ingress and egress edge module 22 is preferably collocated with each core module 34 and serves or hosts a controller, as a controller for the core module, as will be explained below in more detail.

FIG. 4 shows an exemplary distribution of a WDM-TDM packet switch 20 in accordance with the invention, to illustrate a hypothetical geographical distribution of the switch. Cross-connectors 36 and optical links 38 are not shown in FIG. 4 for the sake of clarity. In this example, 16 ingress and egress edge modules 22, 24 numbered 0-15 and four core modules 34 numbered 0-3 are potentially distributed over a large geographical area. As explained above, an ingress edge module 22 is collocated with each core module 34. In this example, ingress edge modules 0-3 are collated with corresponding core modules 0-3. Because the space switches 26 are rote devices with substantially no

computational capacity, they require controllers to perform scheduling allocations and other functions which are described below in more detail. The ingress edge modules 22 include high-speed processors which are capable of performing control functions, or hosting control functions, for the core modules 34. Consequently, an ingress edge module 22 is preferably collocated with each core module 34. The processor of the ingress edge module 22 need not, however, perform the control functions of the core module 34. Rather, it may host, at one of its ports, a processor to perform the control functions of the core module 34. The collocation is also important to enable time coordination in the distributed WDM-TDM packet switch 20, as explained below.

**Time Coordination in the Distributed WDM-TDM Packet Switch**

Time coordination is required between ingress edge modules 22 and core modules 34 if the WDM-TDM packet switch 20 geographically distributed. Time coordination is necessary because of propagation delays between ingress edge modules 22 and the core modules 34. Time coordination is accomplished using a method described in Applicant's above-referenced copending patent application filed April 4, 1999. In accordance with that method, time coordination is accomplished using an exchange of timing packets between the ingress edge modules 22 and the respective edge module controller for core modules 34. At predetermined intervals, each ingress edge module 22 is programmed to send a timing packet to the ingress edge module 22 that serves as a controller for the associated core module 34. For example, ingress

edge module 9 (FIG. 4) at a predetermined interval sends a timing packet to ingress edge module 3 associated with core module 3. On receipt of the timing packet, the ingress edge module 3, which serves as a controller for the core module 3, stamps the packet with a time stamp that indicates its local time. At some convenient time prior to the next predetermined interval, the time stamped packet is returned to the edge module 9. The edge module 9, and each of the other ingress edge modules 0-15, maintains an array of *M* reconfiguration timing circuits where *M* equals the number of core modules 34. The core modules 34 operate independently and reconfigure independently, as will be described below in more detail. Consequently, each ingress edge module 22 must maintain a separate reconfiguration timing circuit coordinated with a local time of an ingress edge module 22 collocated with each core module 34. Without timing coordination, guard times for reconfiguration of the core modules 34 would have to be too long due to the propagation delays between the geographically distributed ingress edge modules 22 and the core modules 34.

For example, in the configuration of the WDM-TDM packet switch 20 shown in FIG. 4, each ingress edge module 22 must maintain an array of four reconfiguration timing circuits respectively coordinated with the local times of ingress edge modules 0-3 collocated with the respective core modules 34. As explained above, in order to maintain time coordination, the ingress edge module 9, at regular predetermined intervals, sends a timing packet to the ingress edge module 0. The timing packet is sent over a communications time slot and received on an ingress port of the ingress edge module 0 dedicated to management functions. The ingress port, on receipt of

the timing packet, time stamps the packet with the time from its local time (timing circuit 0) and queues the timing packet for return to the edge module 9. At some convenient later time before the start of the next timing interval, the timing packet is returned to the ingress edge module 9. On receipt of the timing packet at ingress edge module 9, the ingress edge module 9 uses the time at which the packet was received at ingress edge module 0 (time stamp) in order to coordinate its reconfiguration timing circuit 0 with the local time of ingress edge module 0. Several methods for timing coordination are explained in detail in Applicant's copending Patent Application Serial No. 09/286,431 filed April 6, 1999.

**Packet Transfer Through the WDM-TDM Packet Switch**

Ingress and egress edge modules 22, 24 of the WDM-TDM packet switch 20 operate in packet switching mode. The edge modules 22, 24 are adapted to switch variable sized packets and transfer the packets to subtending sinks in the format in which the packets were received. Switching in the core modules 34 is accomplished in circuit switching mode. The core modules 34 are completely unaware of the content switched and simply switch data blocks. In order to improve resource allocation granularity, the WDM-TDM packet switch 20 switches in both wave division multiplexing (WDM) and time division multiplexing (TDM) modes. Each link 38 (FIG. 3) interconnecting the switched edge modules 22, 24 and the core modules 34 is preferably an optical link carrying WDM data on a number of channels, each channel being one wave length in the WDM optical link 38. Each channel is further divided into a

plurality of discrete time slots, hereinafter referred to simply as "slots". The number of slots in a channel is a matter of design choice. In a preferred embodiment, each channel is divided into 16 time slots. Consequently, the smallest assignable increment of bandwidth is $1/16^{th}$ of the channel capacity. For a 10 gigabit per second (10 Gb/s) channel, the smallest assignable capacity allocation is about 625 megabits per second (625 Mb/s). Connections requiring less capacity are aggregated by class-of-service and quality-of-service in a manner well known in the art. Connections requiring more capacity are allocated multiple slots, as required.

**Admission Control**

The capacity requirement for each connection established through the WDM-TDM packet switch 20 is determined either by a specification received from a subtending source or, preferably, by automated traffic measuring mechanisms based on traffic monitoring and inference. If automated measurement is used, the capacity requirements are expressed as a number of slots for high bandwidth connections. For aggregated traffic, the capacity requirements are measured for a class of service. Regardless of the method used to estimate the capacity requirements, it is the responsibility of the ingress edge modules 22 to quantify the traffic requirements for its traffic load. It is also the responsibility of the ingress edge modules 22 to select a route for each admission request. Route selection is accomplished using connection tables provided by a Network Management System (NMS) (not illustrated) which provides a table of preferred connecting core modules between each ingress and egress edge module.

Admission control may be implemented in a number of ways that are well known in the art, but the concentration of responsibility is at the edge and any ingress edge module 22 receiving an admission request first determines whether free capacity is available on any of the preferred routes through a core module defined in its connection table prior to acceptance.

**Scheduling at the Edge**

At any given time, each ingress edge module 22 has an allocated capacity to each egress edge module 24 expressed as a number of slots. The number of allocated slots depends on a capacity allocation, which may be 0 for certain ingress/egress module pairs. The allocated capacities may be modified at regular reconfiguration intervals which are independently controlled by the controllers of the distributed core modules 34. An ingress edge module 22 accepts new connections based on its current capacity allocation to each egress edge module 24. The controller of each ingress edge module 22 also monitors its ingress queues, which are sorted by egress edge module, as described above, to determine whether a change in capacity allocation is warranted. It is the responsibility of each ingress edge module 22 to determine when resources should be allocated and when resources should be released. However, it is the controllers at the core modules 34 that determine whether a bandwidth allocation request can be granted. Bandwidth release requests are always accepted by the controllers of the core modules 34. The re-allocation of bandwidth and the reconfiguration of the core modules 34 is explained below in more detail.

**I CLAIM:**

1. A high capacity packet switch, comprising:

a plurality of core modules, each of the core modules operating in a circuit switching mode;

a plurality of edge modules connected to subtending packet sources and subtending packet sinks, each of the edge modules operating in a packet switching mode;

wherein the core modules switch payload traffic between the edge modules using wavelength division multiplexing (WDM) and time division multiplexing (TDM).

2. A high capacity packet switch as claimed in claim 1 wherein each core module is a space switch.

3. A high capacity packet switch as claimed in claim 2 wherein each core module is a single-stage electronic rotator switch.

4. A high capacity packet switch as claimed in claim 1 wherein:

each edge module has a plurality of ingress ports, each of the ingress ports having an associated ingress queue; and

an ingress scheduler sorts packets arriving in the ingress queues from the subtending packet sources, the sort being by egress edge module from which the respective packets are to egress from the high capacity packet switch for delivery to the subtending packet sinks.

5. A high capacity packet switch as claimed in claim 4 wherein the ingress scheduler periodically determines a number of packets waiting in the ingress queues for each respective egress edge module and sends a capacity request vector to each of the controllers of the core modules, the capacity request vector sent to a given controller relating only to a group of channels that connect the edge module to the given core module.

6. A high capacity packet switch as claimed in claim 5 wherein each ingress edge module maintains a vector of pointers to the packets sorted by egress edge module and a scheduling matrix that provides a port number for each slot in which a data block can be transferred, the scheduling matrix being arranged in the same egress edge module order so that the scheduling matrix and the pointers are logically aligned; and, when a non-blank entry in the scheduling matrix indicates an egress port through which a data block can be transferred, a corresponding pointer in the vector of pointers is used to locate a starting point for the data block in the packets waiting in the ingress queues.

7. A high capacity packet switch as claimed in claim 1 wherein the core modules and the edge modules are spatially distributed.

8. A high capacity packet switch as claimed in claim 7 wherein one edge module is co-located with each core module, and the edge module serves as a controller for the core module.

9. A high capacity packet switch as claimed in claim 8 wherein each edge module has *M* reconfiguration timing circuits, where *M* is the number of core modules, each of the reconfiguration timing circuits being time-coordinated with a time counter in the respective edge modules that serve as processors for the core modules, to coordinate data transfer from the ingress edge modules when the core modules are reconfigured to change channel connectivity.

10. A high capacity packet switch as claimed in claim 1 wherein each edge module is connected to each core module by at least one communications link.

11. A high capacity packet switch as claimed in claim 10 wherein each core module comprises a plurality of single-stage rotator switches, each rotator switch having a number of input ports collectively adapted to accommodate a number of channels equal to the number of ingress edge modules and a number of output ports collectively adapted to accommodate a number of channels equal to the number of egress edge modules, and each edge module has at least one channel to each of the rotator switches.

12. A high capacity distributed packet switch, comprising:

a plurality of distributed core modules, each core module having a plurality of input ports and a plurality of output ports, the distributed core modules switching payload traffic in a circuit switching mode; and

a plurality of distributed ingress edge modules, each ingress edge module having a plurality of ingress ports for receiving payload traffic from subtending sources and a plurality of egress ports for transferring payload traffic to the core modules;

a plurality of egress edge modules having a plurality of ingress ports for receiving payload traffic from the core modules and a plurality of egress ports for transferring the payload traffic to subtending sinks; and

each of the ingress and egress edge modules operates in a packet switching mode.

13. A high capacity distributed packet switch as claimed in claim 12 wherein the ingress edge modules and the egress edge modules comprise integrated units of one ingress edge module and one egress edge module each.

14. A method of switching payload data packets through a distributed data packet switch, comprising the steps of:

a) receiving a payload data packet from a subtending source at an ingress edge module of the distributed data packets switch;

b) determining an identity of an egress edge module from which the data packet should egress from the distributed data packet switch;

c) arranging the data packet in a sorted order with other data packets received so that the data packets are arranged in a sorted order corresponding to the identity of the edge module from which the data packet should egress from the distributed data packet switch;

d) transferring the sorted data packets in fixed-length data blocks to a core module of the distributed data packet switch;

e) switching the fixed-length data blocks through the core module to the egress module; and

f) transferring the payload data packet from the egress module to a subtending sink.

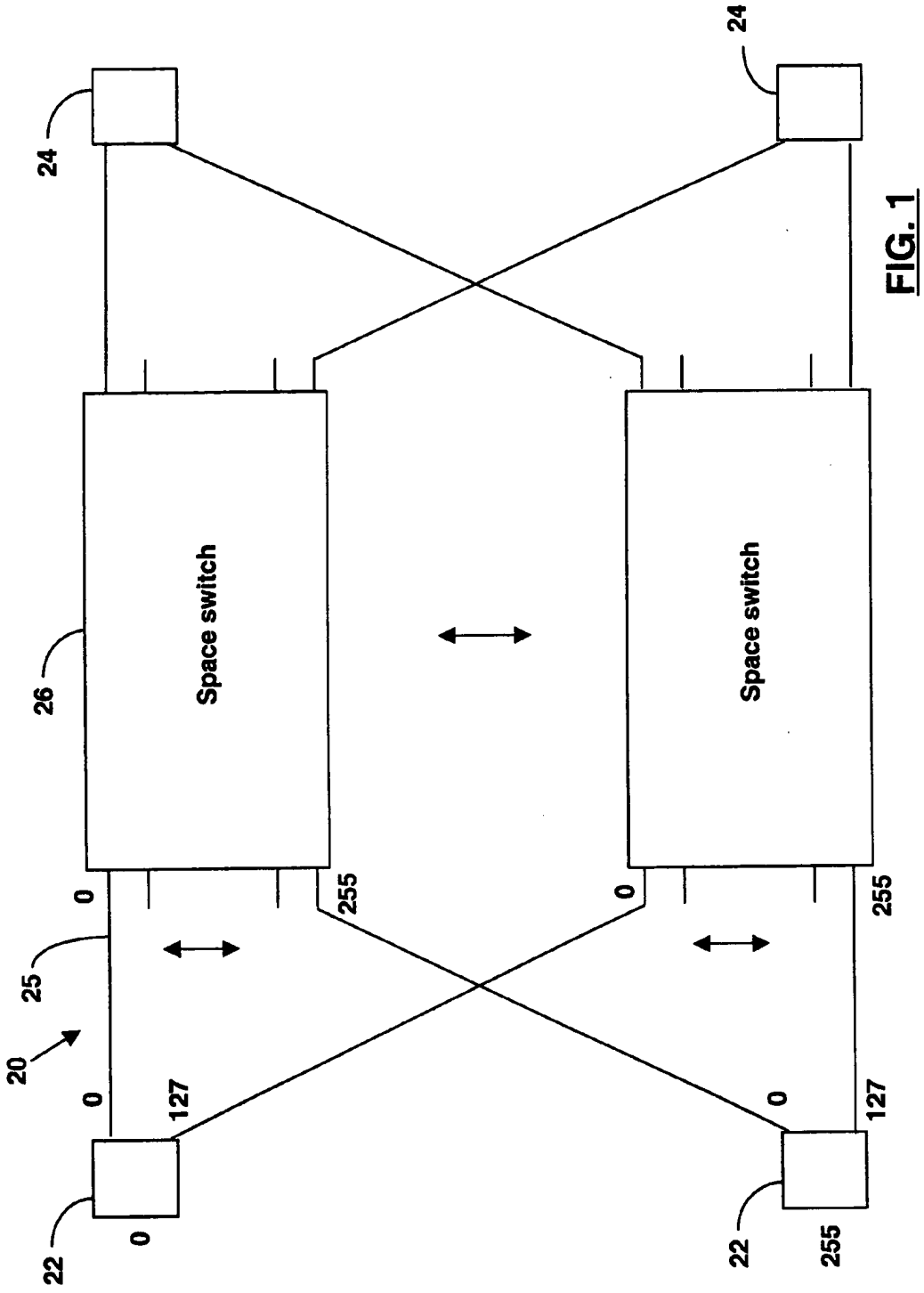
24
24
Space switch
Space switch
26
0
255
0
255
25
20
0
127
0
127
22
0
22
255

FIG. 1

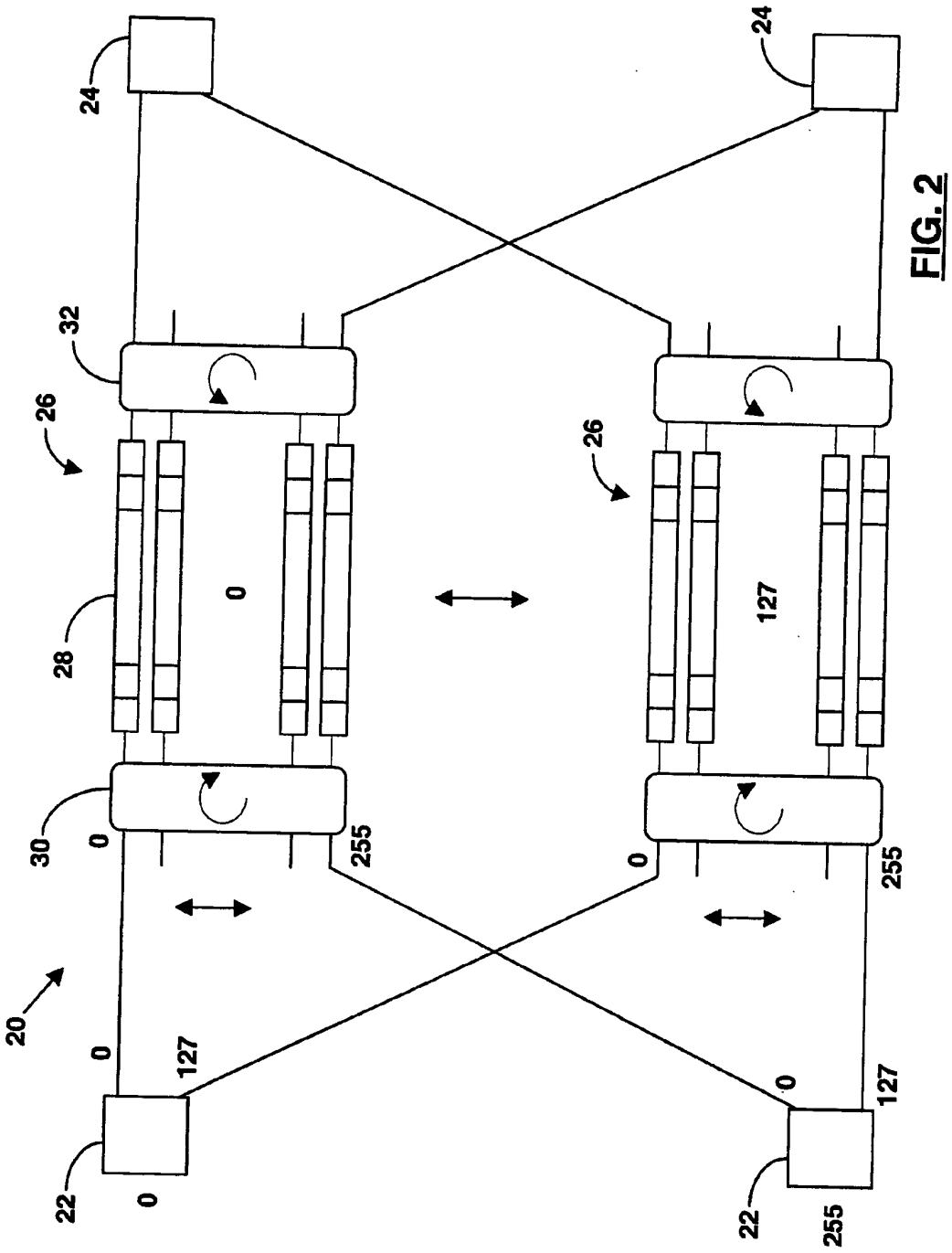


FIG. 2

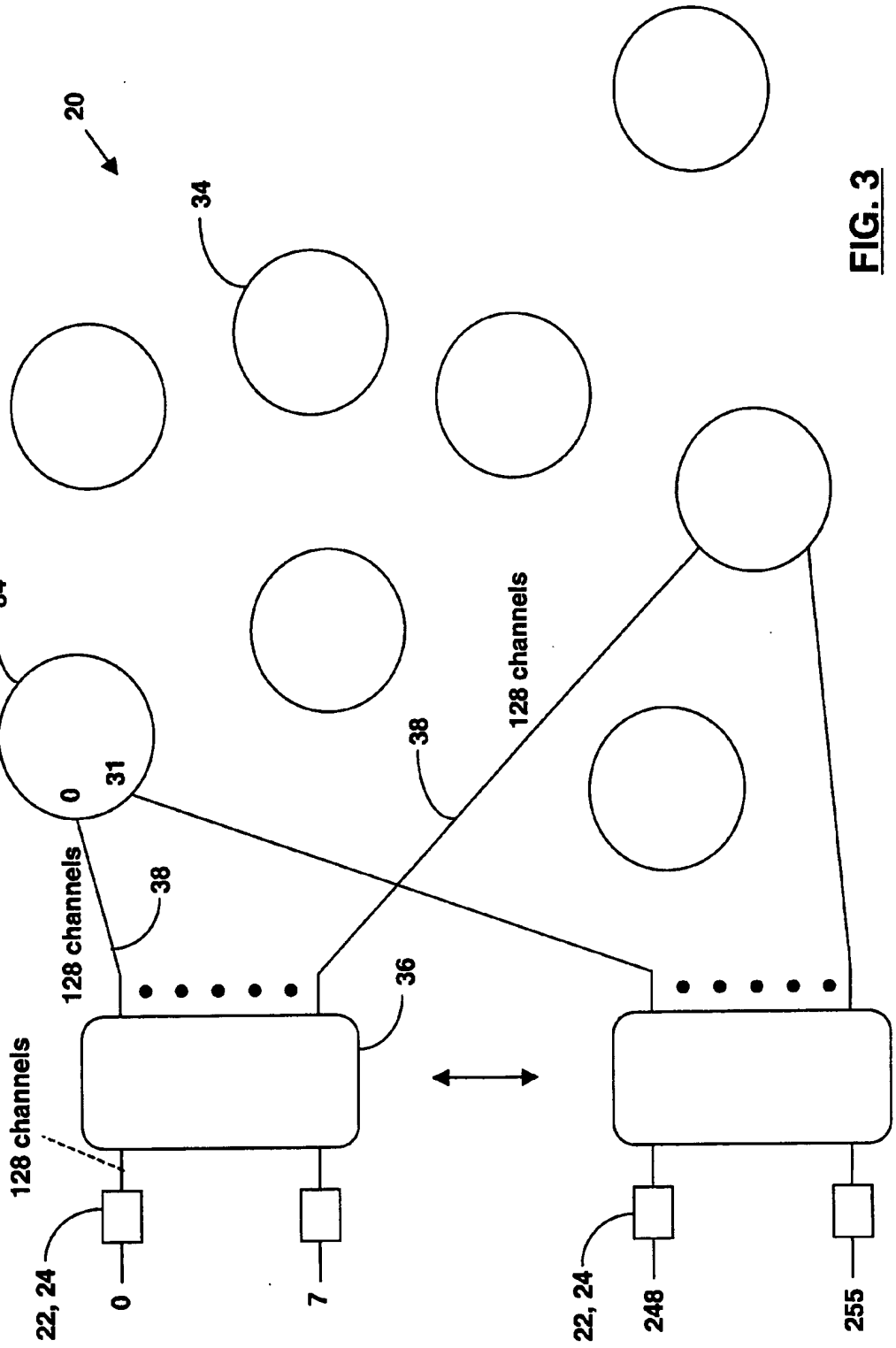
20
34
34
0
31
38
128 channels
128 channels
38
128 channels
36
128 channels
22, 24
0
7
22, 24
248
255

FIG. 3

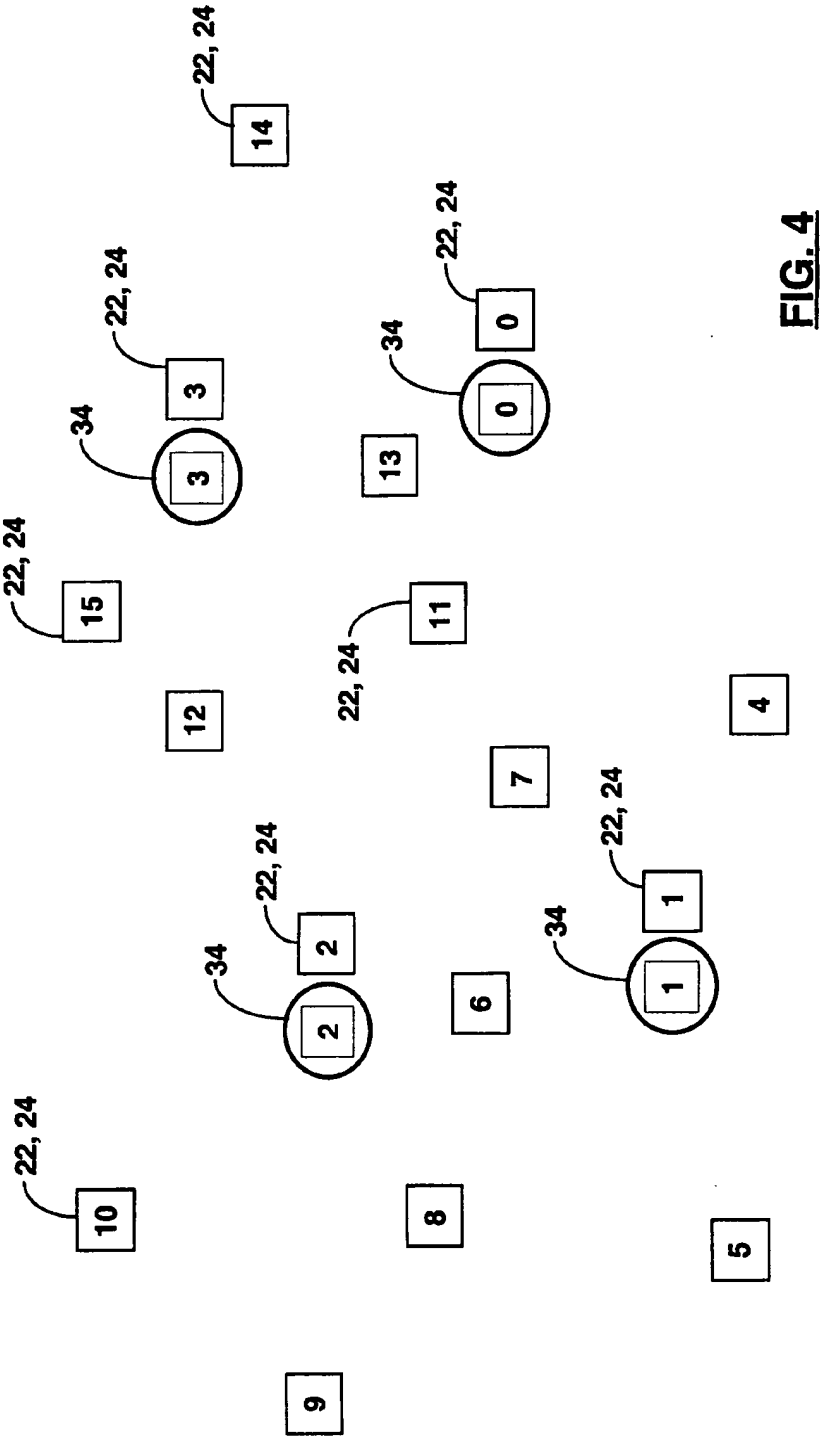
22, 24
14
22, 24
3
34
3
22, 24
0
34
0
13
22, 24
15
22, 24
11
12
7
4
22, 24
2
34
2
22, 24
1
34
1
6
22, 24
10
8
5
9

FIG. 4

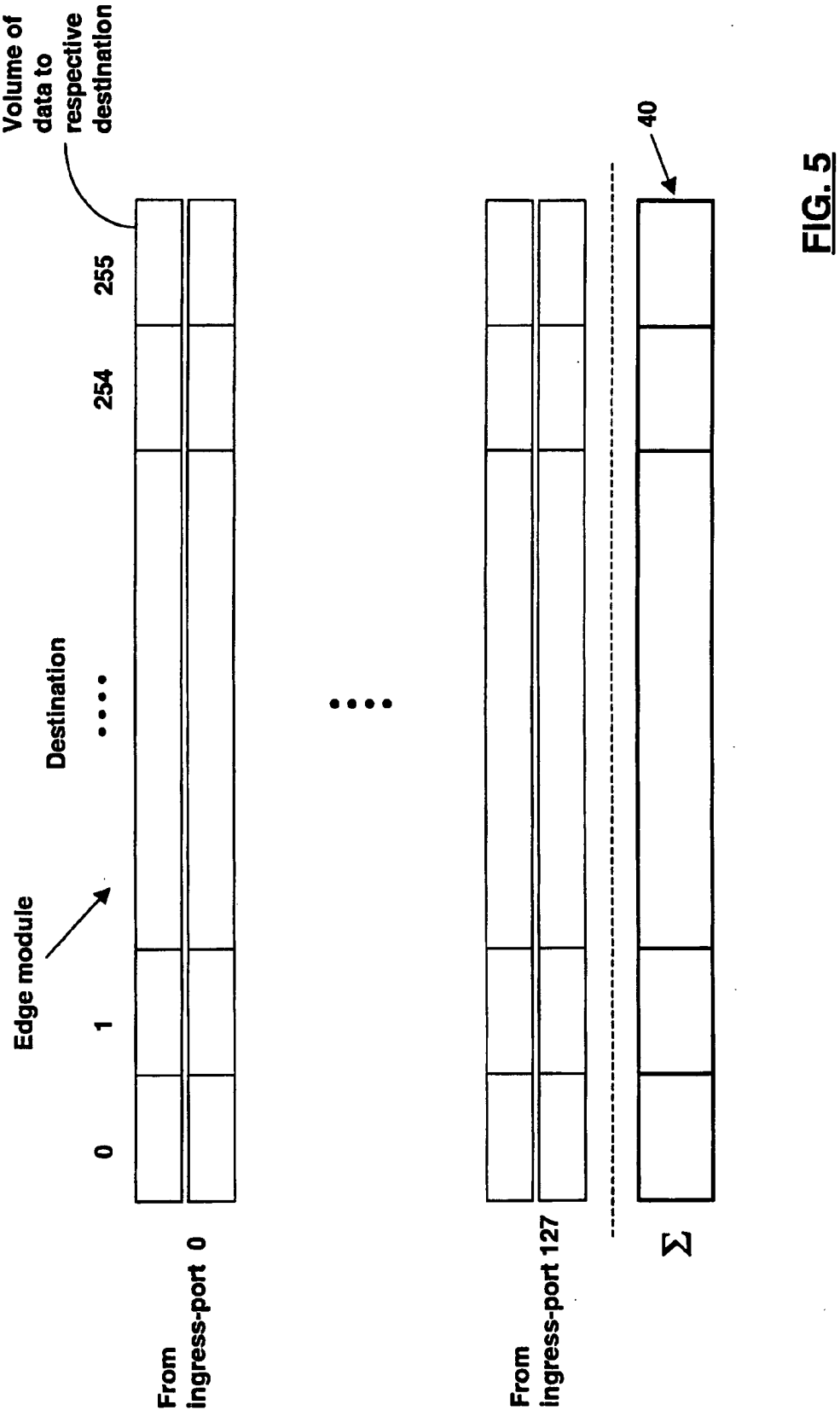
Volume of data to respective destination
Edge module
Destination
0
1
254
255
40
From ingress-port 0
From ingress-port 127
Σ

FIG. 5

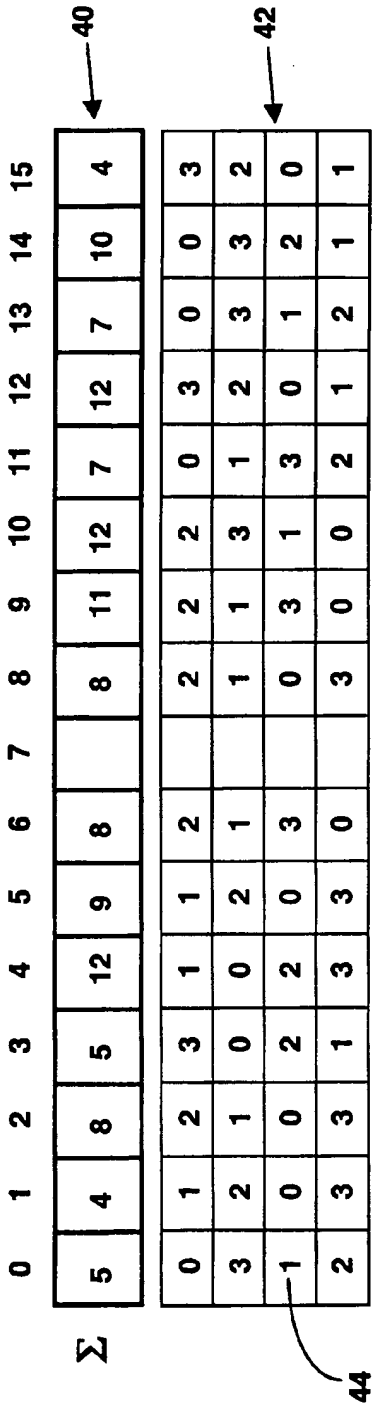


FIG. 6

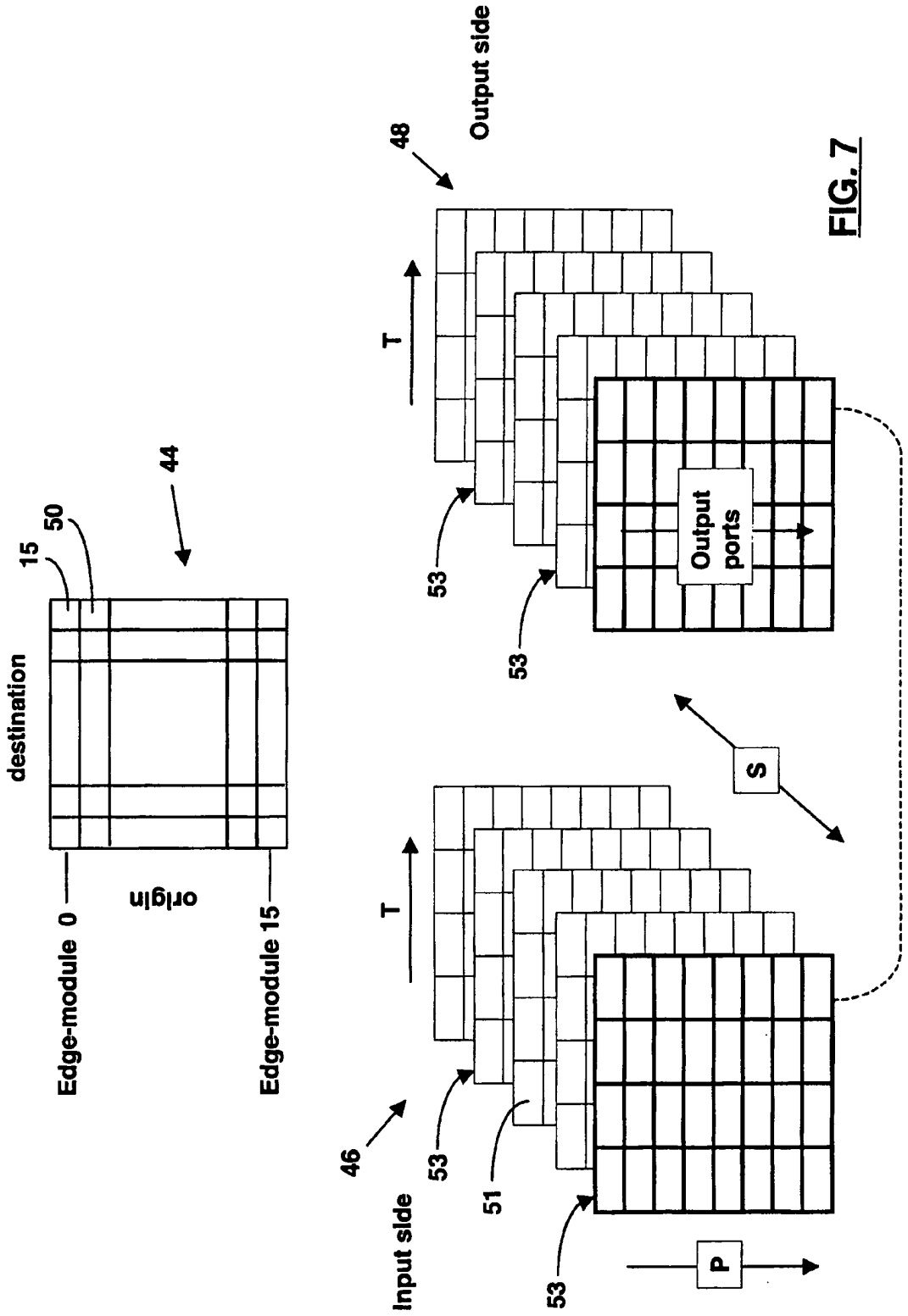
Output side
48
T
Output ports
53
53
S
T
46
53
51
53
Input side
P
44
50
15
destination
origin
Edge-module 0
Edge-module 15

FIG. 7

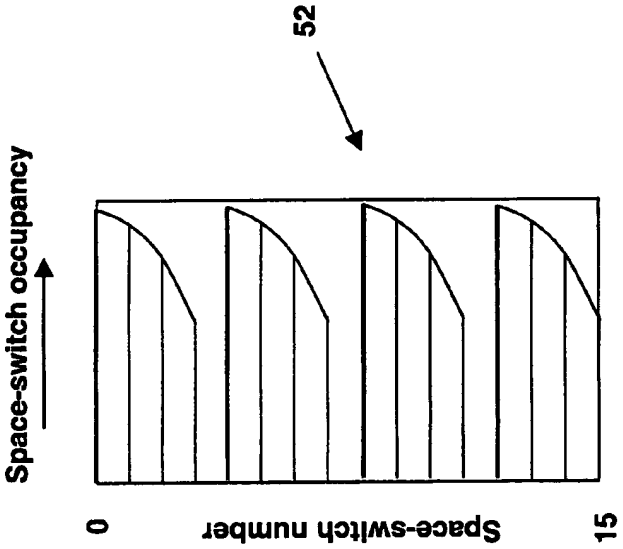


FIG. 8

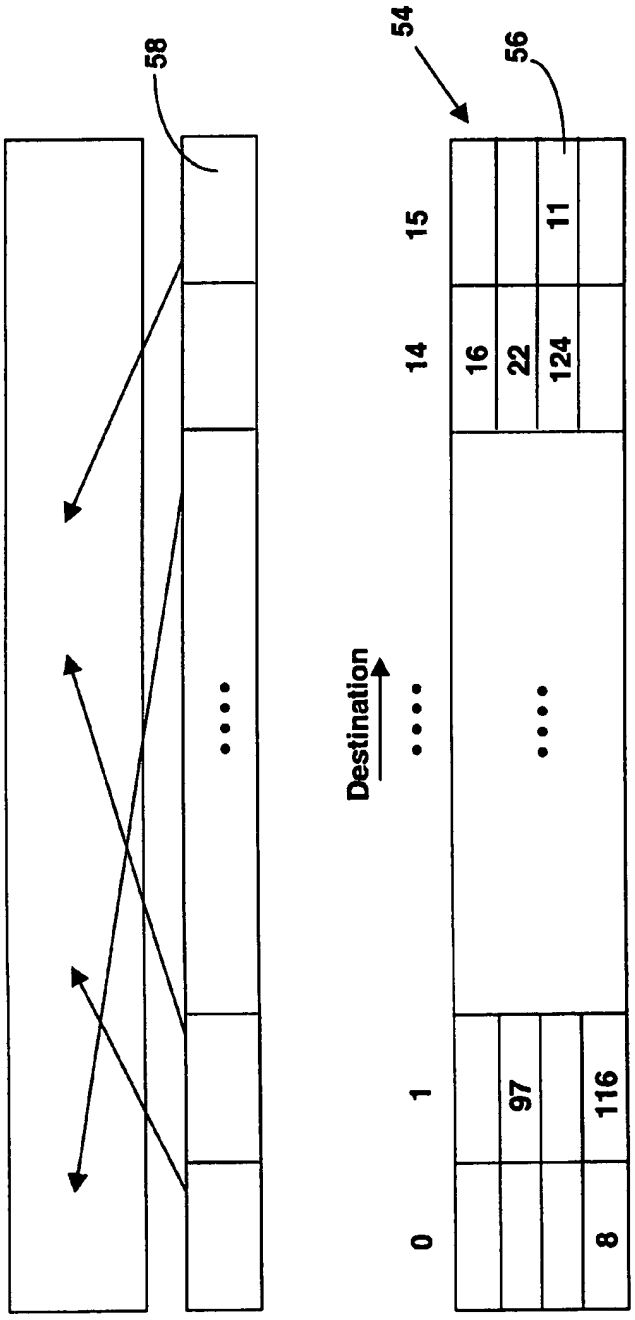


FIG. 9